

```
In [1]: # Dependencies and Setup
import pandas as pd
```

```
In [2]: # Files to Load
school_data_to_load = "../Resources/schools_complete.csv"
student_data_to_load = "../Resources/students_complete.csv"
```

```
In [3]: # Read School and Student Data File and store into Pandas DataFrames
school_data = pd.read_csv(school_data_to_load)
student_data = pd.read_csv(student_data_to_load)
```

```
In [4]: # Combine the data into a single dataset.
school_data_complete = pd.merge(student_data, school_data, how="left", on=["sc
hool_name", "school_name"])
school_data_complete.head()
```

Out[4]:

	Student ID	student_name	gender	grade	school_name	reading_score	math_score	School ID	
0	0	Paul Bradley	M	9th	Huang High School	66	79	0	Dis
1	1	Victor Smith	M	12th	Huang High School	94	61	0	Dis
2	2	Kevin Rodriguez	M	12th	Huang High School	90	60	0	Dis
3	3	Dr. Richard Scott	M	12th	Huang High School	67	58	0	Dis
4	4	Bonnie Ray	F	9th	Huang High School	97	84	0	Dis

```
In [5]: #Check to see if there are any missing values by displaying True or False
school_data_complete.isna()
```

Out[5]:

	Student ID	student_name	gender	grade	school_name	reading_score	math_score	School ID
0	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False
...
39165	False	False	False	False	False	False	False	False
39166	False	False	False	False	False	False	False	False
39167	False	False	False	False	False	False	False	False
39168	False	False	False	False	False	False	False	False
39169	False	False	False	False	False	False	False	False

39170 rows × 11 columns



```
In [10]: #Get all school names
school_names=school_data_complete["school_name"].unique()
print(school_names)
```

```
['Huang High School' 'Figueroa High School' 'Shelton High School'
'Hernandez High School' 'Griffin High School' 'Wilson High School'
'Cabrera High School' 'Bailey High School' 'Holden High School'
'Pena High School' 'Wright High School' 'Rodriguez High School'
'Johnson High School' 'Ford High School' 'Thomas High School']
```

```
In [12]: #Calculate the number of schools
total_number_of_schools=len(school_names)
print(total_number_of_schools)
```

15

```
In [14]: #Calculate the number of students
total_number_of_students=school_data_complete["student_name"].count()
print(total_number_of_students)
```

39170

```
In [68]: #Calculate total budget
filter_school_df=school_data_complete.drop_duplicates(subset='school_name')
total_budget=filter_school_df["budget"].sum()
print(total_budget)
```

24649428

```
In [16]: #Get statistics of the data frame
school_data_complete.describe()
```

Out[16]:

	Student ID	reading_score	math_score	School ID	size	budget
count	39170.000000	39170.000000	39170.000000	39170.000000	39170.000000	3.917000e+04
mean	19584.500000	81.87784	78.985371	6.978172	3332.957110	2.117241e+06
std	11307.549359	10.23958	12.309968	4.444329	1323.914069	8.749987e+05
min	0.000000	63.00000	55.000000	0.000000	427.000000	2.480870e+05
25%	9792.250000	73.00000	69.000000	3.000000	1858.000000	1.081356e+06
50%	19584.500000	82.00000	79.000000	7.000000	2949.000000	1.910635e+06
75%	29376.750000	91.00000	89.000000	11.000000	4635.000000	3.022020e+06
max	39169.000000	99.00000	99.000000	14.000000	4976.000000	3.124928e+06

```
In [17]: #Calculate average math score
average_math_score=school_data_complete["math_score"].mean()
print(average_math_score)
```

78.98537145774827

```
In [18]: #Calculate average reading score
average_reading_score=school_data_complete["reading_score"].mean()
print(average_reading_score)
```

81.87784018381414

```
In [38]: #Calculate number of students with passing math score
names_math=school_data_complete[(school_data_complete['math_score']>=70)]
number_of_students_with_passing_math_score=len(names_math)
print(number_of_students_with_passing_math_score)
```

29370

```
In [39]: #Calculate percentage of students with passing math score
percent_math_70_or_greater=number_of_students_with_passing_math_score/total_number_of_students*100
print(percent_math_70_or_greater)
```

74.9808526933878

```
In [40]: #Calculate number of students with passing reading score
names_reading=school_data_complete[(school_data_complete['reading_score']>=70)]
number_of_students_with_passing_reading_score=len(names_reading)
print(number_of_students_with_passing_reading_score)
```

33610

```
In [41]: #Calculate percentage of students with passing reading score
percent_reading_70_or_greater=number_of_students_with_passing_reading_score/total_number_of_students*100
print(percent_reading_70_or_greater)
```

85.80546336482001

```
In [25]: #Get the names of students with passing math AND reading
names_math_and_reading=school_data_complete[(school_data_complete['math_score']>=70) & (school_data_complete['reading_score']>=70)]
names_math_and_reading.head()
```

Out[25]:

	Student ID	student_name	gender	grade	school_name	reading_score	math_score	School ID	
4	4	Bonnie Ray	F	9th	Huang High School	97	84	0	Dis
5	5	Bryan Miranda	M	9th	Huang High School	94	94	0	Dis
6	6	Sheena Carter	F	11th	Huang High School	82	80	0	Dis
8	8	Michael Roth	M	10th	Huang High School	95	87	0	Dis
9	9	Matthew Greene	M	10th	Huang High School	96	84	0	Dis

```
In [26]: #Calculate the number of students with passing math AND reading
number_math_and_reading=len(names_math_and_reading)
print(number_math_and_reading)
```

25528

```
In [27]: #Calculate the percentage of students passing math AND reading
percent_math_and_reading=number_math_and_reading/total_number_of_students*100
print(percent_math_and_reading)
```

65.17232575950983

```
In [69]: #Create a data frame to hold districy summary details
district_summary_df=pd.DataFrame({"Total Schools":total_number_of_schools,
                                  "Total Students":total_number_of_students,
                                  "Total Budget": total_budget,
                                  "Average Math Score": average_math_score,
                                  "Average Reading Score": average_reading_score,
                                  "% Passing Math": percent_math_70_or_greater,
                                  "% Passing Reading": percent_reading_70_or_greater,
                                  "% Overall Passing": percent_math_and_reading_70_or_greater},index=[0])
district_summary_df
```

Out[69]:

	Total Schools	Total Students	Total Budget	Average Math Score	Average Reading Score	% Passing Math	% Passing Reading	% Overall Passing
0	15	39170	24649428	78.985371	81.87784	74.980853	85.805463	65.172326

In []: